International Journal of Computing, Programming and Database Management

Parvinder Singh
Research Scholar, Department of Computer Science Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India

Dr. Khushboo Bansal
Assistant Professor Department of Computer Science Engineering Desh Bhagat University, Mandi Gobindgarh, Punjab, India

Review on a rule based approach to align natural language with query language

Parvinder Singh and Dr. Khushboo Bansal

Abstract
Information is crucial to our daily life in the globe. Databases are one of the main sources of information. Massive amounts of data are stored in databases. For access to that data, knowledge of structured query language is necessary. The requirement for a system that converts text to a SQL query itself rather than the user arises from the fact that not all non-technical people can do that. Due to its importance in industry and commerce, the text to SQL task is also essential. The system that supports the text-to-SQL job is called Natural Language Interface to Database (NLIDB). The creation of the NLIDB system has long been a challenge. In this paper, we review the current work and summarize the research trends of NLIDB.

Keywords: NLP, Database, natural language interface, SQL

1. Introduction
Natural Language Processing (NLP) is a field of research and application that determines the way computers can be used to understand and manage natural language text or speech to do useful things. Language experts use the term natural to distinguish between human languages (such Punjabi, Hindi, English, and so on) and computer languages (like C, C++, Java, and Prologue). One of the most active fields in human-computer interaction is natural language processing.

NLP aims to make it possible to communicate with computers without having to memories intricate instructions and processes. In other words, NLP is a method for teaching a computer to understand the natural languages that people use. NLP researchers work to learn more about how people interpret and utilize language so that the right tools and methods may be created to help computers comprehend and manipulate natural languages in order to carry out the necessary tasks. Computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology, and other fields form the basis of NLP [2].

Natural Language Interface to Database (NLIDB)
It is one of the applications for natural language processing and works by asking inquiries in natural language to a database in order to discover responses. The main obstacle for any NLP that enables computers to derive meanings from natural language queries is natural language understanding. Keywords and statistical keyword disambiguation utilizing a vector similarity measure will be used to process natural language in order to retrieve query information. We defined keywords as terms or expressions with a specific meaning inside a domain [1].

The creation of a natural language interface to database systems (NLIDB) is one of the most extensive and fascinating areas of natural language processing (NLP). Numerous NLIDB systems have been created during the past few decades. These technologies enable users to connect with databases in a more flexible and convenient manner [3]. We need information to get through every day. Databases are one of the main sources of information. Three different types of elements make up a database: relations, attributes, and values. Each element is different and special; a value element is the value of a specific attribute, while an attribute element is a specific column in a certain relation. A value and the relation that contains its attribute are compatible with one another [4]. The programmer must write code to execute the queries in databases that offer low-level access techniques, such as flat file databases. The underlying access routines are organized by a particular part of the DBMS called the Query

Corresponding Author:
Parvinder Singh
Research Scholar, Department of Computer Science Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India

Processor to meet higher level database query languages like SQL. As a result, questions can be defined in terms of the desired outcomes rather than in terms of how to get them. A database management system is made up of a number of connected programmes that may access these data. Natural Language Interfaces to Databases (NLIDBs) offer a mechanism for those users to access the database with their native language. Database management systems require the use of complex languages like SQL, SPARQL, and others, which are challenging for causal and non-technical users.

In India most working areas that using Hindi language in day to day activities and that records their data to database in Hindi language but there is no knowledge of Query languages like SQL to retrieve their data from database and update and delete their data to database. The goal of this study is to develop Hindi language interface in database in order to show the possibility of Hindi query processing in database using natural language interface. It provides communication between user and computer without knowledge of SQL query or using their native language.

1.2 Sub components of NLIDB
Computer scientists have divided the problem of natural language admission to an information base into two sub-parts [5]:
- Linguistic component
- Database component

1.2.1 Linguistic Component
It must provide a characteristic language interpretation in response to an appropriate question and generate a characteristic language response based on the results of the information base enquiry.

1.2.2 Database Component
It performs standard Database Management duties. The characteristic information expressions are arranged on the appropriate articles (connection names, trait names, etc.) of the database using a dictionary, which is a table. The semantic mediator and the parser both use the vocabulary.

A typical language generator studies the parse tree utilizing the proper reaction as its information in order to produce enough regular language reaction.

Natural language database systems leverage syntactic information as well as information about the actual data set to properly relate a normal language contribution to the structure and content of that information base. Syntactic information often exists in the etymological section of the framework, particularly in the punctuation analyzer, even though information about the actual data set is partially contained in the semantic information model that is being employed.

Questions in everyday language were converted into a statement in the appropriate inquiry language. The information base administration framework manages the enquiry and produces the necessary data when the claim is presented clearly. At that time, when age schedules lead to a surface language form of the response, the common language segment got these details.

2. Literature Review
Hanane Bais et al. [6] developed a non-exclusive regular language interface with an AI technique for a social

information base. The advantage of this interface is that it has a freely usable information base that naturally grows over time as a result of experience.

Saravjeet Kaur et al. proposed [7] a way for querying the knowledge bases employing methods for a distinctive language interface. An ongoing discussion in the field of information base administration is how to provide non-technical clients with a sophisticated interface. Regular customers are not familiar with conventional languages like SQL. The issue then arises from how they connect to the information base framework. It's possible that a normal customer won't be able to handle the framework of the information base. The paper introduces an interface module that transforms a client's query from natural language into an associated SQL order. Posing queries to information bases in simple languages like English is a very practical and clear approach for users who are unfamiliar with complex data set query languages, such as SQL, for information access from data set structure. The engineering for converting an English Query into SQL was put out in this work.

The Indo-Aryan dialect family comprises the following prominent languages, according to Wiqas Ghai et al. [8].
Punjabi, Hindi, Marathi, Gujarati, Sindhi, Bengali, Nepali, Sinhala, Oriya, Assamese, and Urdu. These dialects are mostly spoken in India, Pakistan, Bangladesh, Nepal, Sri Lanka, and the Maldive Islands. These dialects all contain a wide variety of phonetic components. It has been challenging for this technology to stay up with the research being done and being done for the various dialects in the remainder of the world because few professionals have worked on enhancing Automatic Speech Recognition Systems for the majority of these dialects over the previous 20 years.

The tenth most talked language in the world, English, has not seen any significant advancements in the field of programmed discourse acknowledgement. English deserves study in this quickly developing field of automatic speech recognition since it is a language with a strong literary tradition and is a member of the Indo-Aryan dialect family. This study analyses the attempts performed by several analysts to create programmed discourse recognition frameworks for the majority of Indo-Aryan dialects, and then discusses how well they apply to Punjabi so that a strong task can be started for Punjabi.

Simarjeet Kaur et al. [9] claim that giving detailed and useful instructions for the construction of a functional creepy crawly bug is a crucial element in lowering the nuisance on vegetable crops. After the executives' practice sessions in various vegetable harvests, the ranchers will be instructed to take prompt and effective activities for the board in their fields using this online framework, per the advice of Punjab Agricultural University. The data foundation for this framework was designed with MySQL and developed with PHP, HTML, CSS, JavaScript, and Ajax. The proposed framework works well since it is easy to use, practical, and efficient in lowering bug issues by supplying precise and timely data at a reasonable cost.

Ramesh et al. [10] claim that organized information storage is required due to the internet's volume of data in order to facilitate discovering, retrieving, and maintaining information. An information base is a technological advancement that organizes and standardizes information storage. The structured question language (SQL) becomes

essential knowledge for effectively using large data bases. The use of SQL, however, prevents clients without the required information from accessing data bases. An interface is required in order to allow non-master clients to access these knowledge bases. This article outlines the strategy for developing a Telugu-language information base interface. In order to construct a Telugu language interface, a bank information base is used as a contextual inquiry.

The framework's presentation has proven to be appetizing.

According to Ashish Kumar *et al*. [11], data is crucial to registration in our daily lives. One of the primary sources of data is information bases. The development of databases and information bases has a significant impact on how PC usage is increasing. Most IT applications store information or data, then retrieve it from a database.

Database management systems (DBMS) have frequently been used to handle information storage and retrieval. However, because of the inflexible approach to client participation in their interface, information bases are frequently difficult to use. Data from information bases must be stored and retrieved using data set languages, such as SQL. Data stored in information bases can be accessed and managed using the SQL standard, which is set forth by ANSI. However, nobody will be able to write the SQL query because everyone might not be familiar with the information base's data base and SQL's punctuation and syntax independently. In India, Hindi is the language that most people use by default. Information bases are also utilized by several e-administration programmes. Therefore, these programmes need to be able to recognise a simple Hindi sentence and cycle it in order to create a SQL query so that users who are more familiar with Hindi can utilize such information base applications without any problems.

The SQL query is additionally executed on the database to generate the results. Any Hindi-language interface will therefore be helpful to these people. This article examines the planning of the Hindi language SQL query that the client entered.

According to Jaspreet Kaur *et al*. & #39;s research, Question Answering is becoming a more crucial field of study for specialists as technology advances. Customers enquire about a wide range of topics in an effort to obtain accurate answers from Question Answering Systems. Question responding is the ideal course of action when a client asks a question in common parlance as opposed to an inquiry in order to get specific and correct replies. Among the well- known Indian dialects are Bengali, Hindi, and Telugu. A lot of study is being done in these and other Indian dialects, and analysts are now arguing these languages. This study investigates the functionality of Question Answering Systems in several Indian dialects. We go over and talk about Question 1, one of the great modern idioms. Many organizations frequently underuse the general information answering frameworks created in different Indian dialects.

Manu Bansal *et al*. [13] discovered that the term & quot; information mining & quot; was the most well-known at the time. It is necessary to mine these data repositories for information and fascinating examples. The major goal of the flow research is to employ information mining on a library the board framework Information mining frequently focuses on information repositories or distribution hubs. Costs for a range of items, including as programming, hardware, support, and specialists, go up as a result. Focusing on how existing data contained in the data set may be made useable without developing a new information hub is the main objective. The three main focuses are comprehending the complex point of view, contrasting goals, and creating a model for data extraction from the continuous library knowledge base utilizing the ARM (Association Rule Mining) mining technique. Because SQL (Structured Query Language) can be used to mine information rather than utilizing specific information mining calculations, the analysis also contrasts SQL-based mining with ARM. Because of the boundaries (backing and check) utilized in the information mining calculation, which allow ARM to regulate examples to be extracted considerably more successfully than SQL, the results show that affiliation rule mining outperforms SQL-based mining. The calculations are actualized using ARMADA from MATLAB (Matrix Laboratory) and SQL.

According to studies by Ravinder Kumar *et al*. [14], we require data in our daily lives. One of the primary sources of data is information bases. Databases are the primary source of data for almost all applications, so understanding data set dialects like SQL is essential. To write SQL queries, one needs to be conversant with proper question syntax. No one is able to create SQL queries because of this. To get rid of the unpredictability, many research have abandoned SQL in preference of Natural Language (NL) such as English, French, Tamil, Arabic, Hindi, Punjabi, and others. The use of NL (NLIDB) has inspired the creation of a novel handling strategy known as Natural Language Interface to Database frameworks. Numerous research have previously been conducted on how to handle typical languages, but there is still more that can be done to improve and learn about Hindi and other Indian dialects. As a result, we are striving to create a platform and knowledge base that will enable ranchers to react to the various farming-related queries that have been raised in their native Hindi language in order to accomplish this goal.

We may create the machinery that satisfies the requirements of our job by using Java swinging as the front end and MySQL 5.0 as the back end.

According to Priyanka Arora *et al*. [16], database management systems have been widely used for accessing, storing, and retrieving data. But information base frameworks are not suitable for every client because they are difficult to utilise. Many e-administration systems, including those for banks, railroads, bills, farming, and other industries, rely on information bases. Some consumers find it challenging to use these information base frameworks because they are unfamiliar with the dialects that are utilized. Therefore, they need a system that can interpret a Hindi sentence as an inquiry, prepare it, execute it, and output the result in Hindi. The clients would then lack a strong incentive to learn any challenging low-level dialects, like SQL, used in information bases.

According to Rohini B. Kokare *et al*. [15], when a question is asked in everyday language, the normal language question developer interface pulls the necessary data from the data collection. For the client to successfully extract the necessary data from the information source, they should possess sufficient in-depth knowledge of SQL statements. The Characteristic Language Query Builder Interface (NLQBI) will be able to fix this issue. A significant

development in typical language parsing is getting extremely correct syntactic inspection. By parsing common languages, a data string or sentence can be designed to represent its grammatical structure. One of the parsing methods is reliance parsing. Word associations that remove uncertainty are the focal point of reliance parsing. A sizable portion of the new competent computations for reliance parsing are calculated using the reliance trees. Chart-based reliance parsing models are widely employed in reliance parsing because to its condition of workmanship precision and efficacy. An overview of dependence parsing methods and some recent developments in NLQBI frameworks are presented in this study.

In contrast to most client PC interfaces, a different language interface, according to Preeti Verma *et al.* [17], enables users to speak with a PC framework without any prior preparation. Information bases are usually challenging to use to aid clients because of their inflexible interface. Customers should be able to place orders and ask queries in their own language, and a good NLIDB will reply in that language after they have been understood.

For the many applications that involve communication between people and computer systems, it is advantageous to offer a user-friendly interface. Punjabi is the native language of more than 110 million people worldwide, making it simple for them to use a variety of e-administration applications like Punjab Sewa, Suwidha, online public utility forms, online grievance cell, land records management system, inheritance matters, e-District, farming, and so on. Locals in Punjab might benefit from an information base with an interface in Punjabi. Punjabi is ninth from the top among the 6,900 dialects that the United Nations recognizes as existing on a worldwide scale. This paper provides a brief overview of the Natural language interface to information base, including its various components, benefits, limitations, strategies, and methodologies.

3. Conclusion

In this study, we looked at the NLIDB. We have outlined the key findings from the field, highlighted the problems and difficulties, and talked about the solutions that have been suggested. Recent research has been contrasted to highlight accomplished performance and uncover limits associated with it. Additionally mentioned are a few solutions that have been used, together with their advantages and disadvantages. To reduce the problems, machine learning principles and NLP approaches are utilized in various combinations. Despite the fact that the accuracy of NLIDB systems is not high enough for practical use, current work is encouraging enough to anticipate new possibilities.

4. References

1. Garg P, Singh R, amp; Goel MB. A Novel Approach to Implement Natural Language Interface for Database in Punjabi Language. Journal of Xidian University. 2020;14(4):1230-1238.
2. Garg P, Marwaha A, Goel MB. Identification and Classification of Reduplication Words in Punjabi Language. International Journal of Scientific & Technology Research. 2020;9(6):532-537.
3. Ashish Kumar, Kunwar Singh Vaisla, Article May. Natural Language Interface to Databases: Development Techniques, India; c2013.
4. Mahesh Singh, Nikita Bhati. A Noval Hindi language interface for databases. India; c2014.
5. Amandeep Kaur, Parteek Bhatia. Retrieval of Information from Database with Natural Language Interface. Published at NCACCET-10 approved by AICTE, CCET Chandigarh; c2010.
6. Hanane Bais, Mustapha Machkour, Lahcen Koutti, Querying Database using a universal Natural Language Interface Based on Machine Learning, IEEE; c2016.
7. Saravjeet Kaur, Rashmeet Singh Bali. SQL Generation and Execution from Natural Language Processing, International Journal of Computing & Business Research, Proceedings of 'I-Society' at GKU, Talwandi Sabo Bathinda (Punjab); c2012.
8. Wiqas Ghai, Navdeep Singh. Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi A Case Study, International Journal of Soft Computing and Engineering (IJSCE). 2012 Mar, 2(1).
9. Kaur S, Sawhney BK, Kaur S. Design and Development of Software for Insect Pest Management of Vegetable Crops using Web Technology, International Journal of Emerging Science and Engineering, 2013 Sept, 1(11).
10. Ramesh D, Sanampudi SK. Telugu Language Interface to Databases, International Journal of Advanced Research in Computer and Communication Engineering, 2013 Jul, 2(7).
11. Kumar A, Vaisla KS. Hindi Language Interface to Database using Semantic Matching", Oriental Journal of Computer Science & Technology. 2013 Jun;6(2):133-140.
12. Kaur J amp;, Gupta V. Comparative Analysis of Question Answering System in Indian Languages, 128X International Journal of Advanced Research in Computer Science and Software Engineering. 2013 Jul, 3(7).
13. Bansal M, Kaur M. Analysis and Comparison of Data Mining Tools Using Case Study of Library Management System, International Journal of Information and Electronics Engineering. 2013 Sept, 3(5).
14. Kumar R. A Convenient Hindi Language Interface to Database for Agriculture Based Queries, International Indexed & amp; Referred Research Journal; c2014 Nov.
15. Rohini B, Kokare, Kirti H, Wanjale. A Survey of Natural Language Query Builder Interface for Structured Databases using Dependency Parsing, International Journal of Computer Applications. 2014 Dec 18;107(5):9-14.
16. Arora P, Goswami P. An Efficient Hindi Language Interface using Relational Databases, International Journal of Research in Computer and Communication Technology, 2014, 3(6).
17. Verma P, Arora S, Batra K. Punjabi Language Interface to Database: A Brief Review; c2014.